DOCUMENT RESUME

ED 086 367 PS 007 111

AUTHOR Evans, Ellis D.
TITLE Measurement Practices in Early Childhood Education.
PUB DATE [73]
NOTE 65p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Behavioral Science Research; *Early Childhood Education; Ethics; Evaluation; *Literature Reviews; *Measurement; *Psychological Testing; Testing Problems

ABSTRACT

This overview of measurement practices in early childhood education discusses some of the basic issues and concerns of both researchers and practitioners. Although techniques for the measurement of young children's behavior received primary attention, procedures for measuring parent and teacher behavior are also discussed. Trends in measurement practices and their implications for early childhood are noted. A section on special considerations in testing includes academic or cultural bias, impact of testing on students, and ethical considerations in testing. An extensive bibliography is included. (ST)

FILMED FROM BEST AVAILABLE COPY

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 036367

MEASUREMENT PRACTICES IN EARLY CHILDHOOD EDUCATION

Ellis D. Evans
University of Washington

An overview of contemporary measurement practices with a focus on emergent trends and implications for research workers and instructional personnel. Also attempted is a selective critical commentary based upon empirical research related to measurement technique.

TOPIC	PAGE NUMBER
Introduction	1
Trends in Measurement Practice	3
The Measurement of Young Children's Behavior	3
Other Developments in Current Measurement Practice	6
Functions of Measurement in Early Childhood Education	7
Basic Problems in Measurement	8
The What of Measurement	9
The How of Measurement	10
Interviews	15
Systematic Observation	17
Standardized, Objective Measures	22
Teacher-Made Tests	29
Other Measurement Techniques	31
The When of Measurement	33
Selecting and Evaluating Measurement Techniques	34
Special Considerations in Measurement	36
Academic or Cultural Bias	37
General Impact of Testing on Students	39
The Ethical Use of Test Results	42
Summary and Conclusions	45

PS 002111

Measurement Practices in Early Childhood Education

INTRODUCTION

There can be little question that generalizations about child development and behavior, teacher effectiveness, and the worth of educational programs are no better than the data from which they come. Such data are in turn a function of the strategies and techniques used to measure behavior. In this paper current measurement strategies and techniques applicable to early childhood education are selectively examined. This examination includes a look at the basic measurement problems faced by both teacher educators and teachers of young children. Special emphasis is placed on how to measure behavior.

The task of writing this chapter was approached with all the verve of a Coronado searching for the Seven Cities of Gold. It was hoped that somehow the spadework for this task would uncover a wealth of innovative measurement techniques, especially techniques that reflect advances over the many shopworn procedures that have so long dominated educational practice. Like Coronado, and not surprisingly, hopes were frustrated. In this writer's judgment, few if any genuine breakthroughs in behavioral measurement relevant for practitioners have occurred in recent years. To be sure, there have emerged many variations on the traditional themes of testing and observation. Moreover, a staggering proliferation of new, but conventional, measuring instruments has occurred. This is notably true for measures of preschool children's language development and pre-academic skills. But innovatory measurement techniques that are both valid and practical for widespread use in the field are indeed few and far between.

This view is shared by others who have recently dealt with the role of measurement in the evaluation of early childhood programs. For example, Kamii

and Elliot (1971) have called for the development of measurement techniques better to match the program objectives of new curricula for young children. These authorities are especially critical of the use of conventional, standardized tests of intelligence, visual perception, vocabulary, and psycholinguistic abilities for the summative evaluation of early childhood programs. The main reason for such criticism is that such instruments simply are not designed for this purpose. Rather, instruments of this kind are usually constructed to classify children, diagnose possible learning disorders, or predict subsequent learning and development.

This problem applies also to procedures for comparing various instructional programs for young children. For example, Butler (1970) has remarked during a review of research about early intervention programs:

> "Instrumentation is a particular problem. What kind of instrumentation is valid if one wishes to compare the outcomes of a cognitive, direct-instruction program with a much more broadly based, informally organized program? What can changes in IQ indicate about the outcomes of these programs when other aspects are not measured?" (Butler, 1970, p. 18).

Some readers will surely quarrel with the pessimistic tone of these conclusions. Therefore, possible exceptions to the general situation should be mentioned. For example, considerable advances in the technique of computer-assisted branched testing have been made (Holtzman, 1971). However, the application of this technique depends upon an elaborate and expensive set of hardware and technical know-how seldom found in school settings. Perhaps even more exciting is the potential for unobtrusive, or inconspicuous, measures in various early childhood education endeavors (Webb et al. 1966). To date, however, this potential has not yet been fully explored.

Trends in Measurement Practice

Lest the foregoing be taken as a too discouraging perception of the measurement field, several encouraging trends in measurement practice can be noted.[1] These trends largely involve developments in the measurement of children's behavior, although a few concern more directly certain other variables such as curriculum components and institutional change. Consider first those trends specific to the measurement of young children's behavior.

The Measurement of Young Children's Behavior

Under this topic, at least seven trends can be identified. First, the range of measures available for use with young children has increased rapidly in the past several years. For example, no longer is a practitioner concerned with preschoolers limited to the use of intelligence scales, developmental "schedules," and highly experimental measures of learning ability. (See Appendix for a listing of recently published tests and scales designed for use with infants, preschoolers, and early school age children.) Especially notable is the move toward comprehensive assessment of children's language development, until recently, and except for vocabulary development, an area sadly neglected in many early childhood education programs. See Cazden (1971) for an overview of procedures for measuring young children's language development.

Second, and related to the first, is a growing concern for the measurement of children's affect, including motives, attitudes, and self-esteem. This trend is reflected in several ways, including the widespread belief among many early childhood educators that the cultivation of a child's affective life is as

---

[1] For an overview of general trends in educational and psychological testing, see the Review of Educational Research, 1965, v. 35, No. 1, and Deal and Wood (1968).

important, if not more so, than is the cultivation of his intellect. An increasing number of early education programs include explicit goals for affective development. Thus there has emerged a need for suitable measures of this aspect of children's behavior. Research about constructs such as intellectual achievement responsibility (Crandall, Katkovsky, and Crandall, 1965), achievement motivation (Heckhausen, 1967), and self esteem (Coopersmith, 1967) exemplifies the dynamic interaction of affective and cognitive developmental factors in children.

A third trend is a growing awareness and respect among measurement specialists and educators for assessing individual differences along cultural-linguistic lines. This trend, described elsewhere as a decrease in the ethnocentrism of psychological assessment (Holtzman, 1971), is perhaps most apparent in the measurement of scholastic aptitude and language competence. For example, some authorities (e.g., Baratz, 1969) are pressing strongly for the construction and administration of tests in the dialect or native language of minority group children. Frequently, such authorities also recommend that test content be altered better to reflect the cultural background of minorities. The educational advantages of such a move have yet to be fully explicated on empirical grounds. However, this approach stands in marked contrast to the practice of using for all ethnic minority children measures of language and academic skills valued by white, middle class adults (including psychologists). The general isuue of cultural bias in testing is considered more fully later in this chapter.

Fourth, there currently appears to be less emphasis on the use of formal tests alone for measuring the behavior of children (and teachers), and a greater emphasis on the use of other techniques, including systematic observation (McReynolds, 1968). This is clearly reflected in the growing popularity of process observation procedures, such as interaction analysis and microteaching. It is also indicated by the focus on children's products (e.g., stories, scientific experiments, art work, and other creative outputs) by protagonists of "open education." These notions are considered in more detail in the section of this chapter concerned with teacher made tests.

A fifth trend concerns the development of systematic procedures for screening preschool and kindergarten-entry children in order better to tailor their preacademic and early academic experiences. These procedures, typically developed at the local school district level in relation to specific programs, stand in contrast to the traditional, global "readiness" test approach or informal teacher ratings of developmental status. Aside from the general purpose of better educational planning for school beginners, screening measurements often provide data for the early detection of learning and behavior disorders. Examples of screening practices based upon the formal application of tests and scales include Ahr (1967), Conrad and Tobiessen (1967), and Rea and Reys (1970). Rogolsky (1969) has provided a brief review of developments in this area. The value of early screening is also reflected in the creation of new tests of children's competence with concepts deemed basic to early school success (e.g., Boehm, 1969; Moss, 1970).

Sixth, there has been a marked increase in infant assessment during the past decade, especially in relation to infant stimulation studies (e.g., Painter, 1968; White, 1971). This interest indicates both a renewed concern for the diagnosis of early developmental status and a bias in developmental theory regarding the importance of early experiences for overall development (Stedman, 1966). Instruments for infant assessment are reviewed by Thomas (1970). These are largely concerned with early cognitive and sensory-motor behaviors. Practically no useful measures of early affective development (birth-age 3) are reported in the recent literature about infancy.

Finally, a growing number of resource books and services relevant to measurement in childhood education are appearing. Hopefully, this means that educators are becoming more aware both of the need for and value of judicious measurement practices in their work with young children. Examples include Beatty (1969), Bloom

PS 004111

measurement practices in their work with young children. Examples include Beatty (1969), Bloom et al (1971), Goolsby and Darby (1969), Hess et al (1967), Jenkins et al (1966), Johnson and Bommarito (1971), McReynolds (1968), Palmer (1970), and Savage (1968). To these resources can be added information services such as the Test Collection Bulletin published regularly by Educational Testing Service. Interested readers are encouraged to consult the aforementioned references and subscribe to the ETS Bulletin. In addition, a publication devoted to the description and evaluation of tests keyed to the objectives of elementary school education (Grades 1-6) is currently available (CSE, 1970).

Other Developments in Current Measurement Practice

In addition to the foregoing trends, at least two other developments in current early education measurement practice can be cited. The first of these involves the search for measurements of program variables other than those expressed solely in terms of pupil or teacher behavior. Three examples can serve to illustrate this trend. First, a technique has been developed for assessing the organization of physical space within which early education occurs (Kritchevsky and Prescott, 1969). This seems particularly useful in view of the apparently significant, but often overlooked, relationship between spatial organization and the classroom-playground behavior of children and their teachers (Prescott and Jones, 1967). Related to this technique are still more comprehensive attempts to measure educational environments, some of which can be adapted to the concerns of teacher educators (e.g., Astin and Holland, 1961; Creager and Astin, 1965), and measurement guidelines for the evaluation of a total school system (MGS, 1964).

Another example of the aforementioned development is the conceptualization of measurement criteria for purposes of evaluating instructional materials and

equipment. Notable among specific developments along this avenue of measurement are procedures designed to measure the reading difficulty or "readability" of written materials (Bormuth, 1968; Klare, 1963). Other more preliminary efforts in this direction appear promising (e.g., Dick, 1968; Eash, 1969) although not yet widely applied in the specific context of early childhood education.

A final example involves the analysis of measurable curriculum dimensions (e.g., pacing, variety, sequencing, and scope) upon which early childhood education programs can be compared (Lay and Dopyera, 1971). The application of this measurement concept to early education research is but in the embryonic stage. However, it seems especially suitable for objectified assessments of diverse instructional programs.

A second, broad development in the measurement of variables other than classroom behavior is represented by attempts to measure institutional change. This is notably the case within communities that are served by programs such as Project Head Start (Kirschner Associates, 1970). Among the quantitative institutional variables amenable to measurement include involvement of the poor in community decision-making activities, employment of local residents in paraprofessional occupations, and allocation of resources to the educational and health needs of poverty and minority groups. This approach to the measurement of change is significant if for no other reason that it encourages one to assess changes that extend beyond immediate pupil outcomes to possible broad scale social benefits of early education programs.

Functions of Measurement in Early Childhood Education

[redacted] Evaluation pervades virtually every aspect of early childhood education, including the preparation of teacher personnel. If evaluation is to be based on data, measurements of one sort or another are

necessary at various points in any program, whether its focus be personnel training or children and their parents. Points for measurement are both frequent and critical. For example, important measurement functions at the preservice level of teacher education often include selection of trainees, diagnosis of trainee needs, measuring trainee progress and training outcomes, and predicting in-service teaching success. At the latter level of training, measurement again becomes central in matters of selecting personnel and determining teacher effectiveness.

For children involved in early education programs, measurement is equally important. If the readiness principle is to be anything but a sterile cliche, children's entering behavior along multiple dimensions must be assessed and the data then used to facilitate individualized instruction where necessary. Many educators are interested in charting the developmental progress of children apart from specific curricular experiences and, of course, the degree of children's progress in relation to explicit curriculum objectives must be measured in some way.

Measurement is also crucial for parents variously involved in early education programs. Increased concern is being shown for measuring the quality and extent of parent involvement in early education, the outcomes of attempts to provide parent education in matters of child development and family relations, and parent satisfaction with their children's participation and progress in given early education programs.

BASIC PROBLEMS IN MEASUREMENT

Most simply, measurement is the description of data in terms of numbers (Guilford, 1954). More specifically, measurement involves the assignment of

numerals to objects or events according to certain rules in order to represent magnitude (Stevens, 1951). Occasionally, one is interested in measurement only to determine the presence or absence of some property, without further quantification in terms of "more" or "less" (English and English, 1958). In this paper the complexities of rules for assigning numbers, scaling procedures, and the like are not considered. Interested readers may review such topics elsewhere (e.g., Stevens, 1951; Nunnally, 1964). It is important, however, to point out that any use of measurement in educational settings involves at least three assumptions: The behavior of children and teachers can be symbolized numerically, the numerical description of behavior can be analyzed according to certain mathematical principles, and the results of such analyses can serve as useful and valid indications of the behavior involved (DuBois and Mayo, 1970). Once these assumptions are accepted, measurement can proceed. But at least three basic problems must be solved by anyone concerned with measuring behavior in an educational program for children or teachers. These are the problems of what, how, and when to measure (Webb, 1970).

The What of Measurement

The effectiveness of behavioral measurement in training programs is contingent upon the precision with which training outcomes are specified. That is, until one decides exactly what it is that a child (or teacher) should be doing differently as a result of a training experience, it is unlikely that measurement will be useful for one's intended purposes. This is faced the problem of establishing objectives.

As Webb (1970) observes, at least two weaknesses often characterize this level of the measurement problem in practice. First, objectives frequently border on the intangible, making difficult any consensus about what constitutes

evidence of the desired behavior. Consider, for example, the ambiguities involved in such kindergarten objectives as "Responsiveness to beauty in all forms" or "Realization of individuality and creative propensities" (Headley, 1965). Second, and related to the first,-is the frequent tendency to determine what to measure on the basis of expediency or convenience. That is, instead of gearing one's measurement policy to relevant program objectives, one opts for measuring what can easily or readily be measured. In the extreme case, one may refrain entirely by not measuring at all on the grounds that suitable techniques are not available, or that the "really important goals" are long term, and therefore measurement at this time is inappropriate.

A measurement policy designed exclusively around program objectives may, of course, be too delimiting. The broader guideline for determining what to measure concerns any information that is either necessary or useful in (1) making decisions about programs and their participants; (2) reporting to outside agencies, parents, and fellow professionals; and (3) charting developmental changes for record-keeping purposes (e.g., height and weight in young children). Hopefully, a measurement policy is never based solely on custom or simply because it is the "thing to do."

The How of Measurement

Once decisions about what to measure have been made, one is faced with the two-pronged measurement techniques problem: (1) determining the units of measurement that are most pertinent to tasks for which an individual is being groomed, and (2) selecting or developing a technique which will yield valid and reliable measurements (Webb, 1970). Commonly used measurement units range from speed, amount, frequency, and accuracy to variety, quality, persistence and originality. For example, a prospective teacher being trained in the successful application of classroom

management techniques may be required to demonstrate that she is capable of handling a child's aggression quickly (speed) in several different ways (variety) that are based upon valid management principles (accuracy). Similarly, in the case of a child being schooled in the techniques of creative problem solving techniques, both persistence and originality, among others, are relevant units of measurement. Measurement of the acquisition of factual knowledge or extent of comprehension of concepts and principles obviously calls for attention to both accuracy and amount. The point is that the unit(s) to be used depend upon the components of behavior that are focussed on in-training. Again, this requires a careful analysis of the behavioral components reflected in program objectives.

The second prong of the problem about how to measure concerns specifically the matter of measurement strategy and technique. By measurement strategy is meant the method for determining the referents against which an individual's behavior can be measured. Measurement technique refers to the particular procedure for describing the behavior, usually in quantitative terms.

Basic Measurement Strategies

Perhaps the most basic distinction in measurement strategy is that between norm-referenced and criterion-referenced measurement (Glaser, 1963) (Popham and Husek, 1969). A norm-referenced measure is one in which the meaning of an individual's behavior is derived from the behavior of others on the same measure. In other words, a comparison of persons whose behavior is measured by the same device is usually necessary for an interpretation of the behavior. The widely used Preschool Inventory (Caldwell, 1967) is an example of this approach to measurement. It is based upon the assumption that individual differences in intellectual attainments exist among children ages $4\frac{1}{2}$ to $6\frac{1}{2}$. A child's

performance on this test is interpreted by comparing him with other children of the same age and socioeconomic standing. Thus, a child's performance can be described as "average," "above average," or "below average" in relation to how the scores of his comparison group are distributed. The distribution of test scores or other quantitative data constitutes the basis for test norms: Hence the term norm-referenced measurement. Most standardized measures of intelligence, academic achievement, and even "personality" are norm-referenced measures.

In contrast, criterion-referenced measurement involves determining an individual's status in relation to some preselected or established standard of performance. This standard (or criterion) -- not other individuals -- becomes the item against which performance is measured and interpreted. Performance tests such as those involved in obtaining a driver's license and demonstrating swimming proficiency are examples. Minimal, although absolute standards of competence must be domonstrated in order to "pass." Insofar as a given individual is concerned, the performance of others on the same measures is irrelevant.

In early childhood education circles, the Basic Concept Inventory (Engelmann, 1967) is an example of a criterion-referenced measure. This measure is based on the assumption that certain basic conceptual skills are critical for successful early academic progress. It can therefore be used to measure which of these various skills the child has or has not mastered so that remedial instruction can be programmed. Or, it can be used to measure the effectiveness of an instructional program designed to develop mastery of these skills among young children.

Among the most recent and comprehensive applications of criterion-referenced achievement measures is represented by the Individually Prescribed Instruction (IPI) evaluation program (Lindvall and Cox, 1970). Developed at the University of Pittsburgh, this program includes four main components: (1) tests for the initial placement of pupils in the instructional program, (2) pretests in relation to specific curriculum unit objectives, (3) curriculum-embedded tests to measure individual pupil progress, and (4) curriculum unit post-tests for summative evaluation. Additionally, non-test information, indluding data obtained during personalized pupil-teacher conferences, are used to facilitate the design of individualized instruction and its evaluation.

The relative merits or weaknesses of criterion- and norm-referenced measurement strategies are perhaps incidental to the basis or rationale for choosing one or the other for use in the

practical setting.[2] In both cases, this choice of strategy is contingent upon the kinds of decisions one will make from the measures obtained (Garvin, 1970). Some educational decisions involve the selection of a "fixed quota" from either the high or low end of a distribution of scores. For example, teacher trainers may wish to admit only those candidates for training whose scores on measures of academic competence and attitude toward children fall in the upper quartile. Or, one may have room for a small number of children in a compensatory education program and select for the special treatment only those who score at some point "below average." In both of these examples, norm-referenced measures would be appropriate. In addition, where information about the capacity of a given instructional program to increase the range of individual differences is sought, norm-referenced measurement is also appropriate.

If, on the other hand, one's decisions are primarily oriented toward certifying competence with respect to some a priori standard, then criterion-referenced measurement is clearly indicated. Training programs where objectives are behaviorally defined are those in which criterion-referenced measurement is natural. In such cases, one is usually most concerned with whether (or what proportion of) students master a given objective, not how they compare to some norm group (Sjogren, 1970).

It should also be noted that norm-referenced measures are often used as if they were measures of the criterion-reference type. A case in point is the use of a conventional intelligence test to "evaluate" the effectiveness of an early intervention program after children have been taught the test items directly. Yet norm-referenced measures typically are designed to "spread out" individuals

---

[2] For a discussion of the merits and limitations of criterion-referenced measurement see Ebel (1970).

along some dimension of behavior, and they usually represent only very broad samples of such behavior at that. Rarely can one find a norm-referenced measure that reflects in specific ways the objectives of most training programs. This point will again be considered, but in a different light, later in the chapter.

Finally, and in relation especially to criterion-referenced testing, increased attention is now being given to the formulation of various rate measures of learning. By this is meant measurements based on time needed by learners to achieve specified goals (Carroll, 1970). A consideration of rate measures involves a number of interacting variables -- motivation (perserverance) of the learner, opportunity to learn, quality of instruction, and learner ability to comprehend and profit from instruction. These variables pose substantial measurement problems in themselves. But, essentially, this approach concerns measuring learning rate in terms of a ratio between amount of knowledge or skill gained and a specified unit of time. Interested readers should consult Carroll (1970) for details.

Measurement Techniques

Measurement techniques can be classified in many different ways: According to the content or area measured (e.g., intelligence, interest, motor skill), the way in which a measure is administered (e.g., group versus individual), response mode (e.g., paper and pencil, free versus controlled response, verbal or nonverbal response), scoring method (e.g., subjective and judgmental versus "objective"), target population (e.g., infants, preschoolers, or teachers), and format of the measure (e.g., rating scale, performance test, process observation). It is convenient, however, to conceptualize measurement techniques along a broad dimension that transcends the foregoing classification schemes, namely, obtrusive-unobtrusive measurements (Shalock, 1968). By definition, an obtrusive measure is one in which an examiner or observer is present on the scene and the examinee or

observee is aware that his behavior is being scrutinized. In general, unobtrusive measures represent the other side of this coin: Physical traces (erosion and accretion), running records, episodic and private records, and the like. To date these measures have been little used in education, although they are potentially valuable, especially in combination with appropriate obtrusive measures (see Webb et al, 1966). The present discussion generally is limited to obtrusive measures. However, it should be recognized that many observational procedures (both simple and contrived) herein discussed are essentially unobtrusive. Moreover, it is even possible that tests, to the extent they become simply another part of classroom routine, may take on the characteristics of an 'unobtrusive' measure.

Obtrusive measures can be grouped into at least five broad categories: Interviews, systematic observation, standardized objective measures, standard projective measures, and teacher-made tests. This writer makes the assumption that readers are sufficiently familiar with these classes of measures (including their principal strengths and weaknesses) that extended descriptions of them are unnecessary. Therefore, only very general comments concerning such classes will be advanced. The thrust instead will be in the direction of highlighting those measures that are currently being profitably used in the field or that seem to hold more than average promise for use by practitioners. This survey is not limited to measures of children's behavior. Attention is also given to examples of measures of teacher and parent behavior.

Interviews. Very few reports involving the use of interview measurements with children in early childhood education programs exist in the current literature. The clearest exception to this is the liberal use of Piaget's methode clinique in curricula based on cognitive developmental theory (Kamii, 1971; Lavatelli, 1970). Such exploratory interviews are semi-structured in that they are designed

within broad limits to determine the child's understanding of observed phenomena related to logical classification, seriation, numerical construction, conservation, and spatial concepts. According to Kamii and Peper (1969) the methode clinique differs from psychometric methods thusly:

> "In the psychometric method, the examiner is required to follow a standard set of procedures specified in the manual, without any deviation. The wording of a question cannot be changed, and the number of times the instruction can be repeated is specified. In the 'exploratory method,' on the other hand, the examiner has an outline and a hypothesis in mind at all times, and he tests these hypotheses by following the child's train of thought in a natural, conversational way. The examiner uses his ingenuity to make himself understood by the child in any way possible." (Kamii and Peper, 1969, p. 13).

For educators interested in the child's conceptualization of Piaget-based tasks, this technique has the potential of yielding information not accessible in any other way. However, this comment is based on the dual assumption that an examiner will execute the methode clinique correctly and will not be deceived by the child's language or his own biases.

Aside from this application, only a scattering of reports of formal interview measurement is apparent in the literature. Perhaps the most novel of these reports concerns the development of a standardized telephone interview procedure for obtaining speech samples from young children. Especially promising results from the use of this technique with disadvantaged children have been reported (IDS, 1968).

Interview measurement can also be helpful in the study of teacher behavior. Currently, however, the interview is largely restricted to two program aspects: selection and program evaluation. The interview has long been a popular procedure for selecting both candidates for teacher training and hiring teachers for existing school programs. One is forced to look long and hard, however, for evidence to indicate that interview data alone predict success in either venture.

Undoubtedly this is due in large part to the difficulty of finding anything in common between interview measurements and in-process teacher behavior. Periodic interviews with teachers-in-training can be useful as a feedback mechanism for training program effectiveness, although surprisingly few examples of this can be found in the current teacher education literature.

Perhaps the most extensive use of interviewing has been in the study of parental child rearing practices. For example, this method has been used profitably in recent years to measure parents' beliefs and perceptions about themselves as parents, including such things as child rearing philosophy and preferred disciplinary practices (Baumrind, 1968). Within the context of early education programs, elaborate and promising interview methods have also been included in broad scale evaluations of Head Start and Project Follow Through such as those executed by agencies like Stanford Research Institute and the Educational Testing Service.

Systematic Observation. Observational techniques need not always be obtrusive. For example, observation conducted through one-way mirrors or by way of video tape recording for later analysis is unlikely to distract or otherwise affect the behavior of children or teachers being observed. However, outside of the psychological laboratory or child-research nursery school, such devices are rarely available (or used). Observation more typically occurs in the presence of children and teachers. Even then systematic, direct observation methods are problematical, especially with respect to reliability and the control of situational variables. Nevertheless, it is this writer's opinion that some of the most promising recent advances in measurement technique have been made in the area of systematic observation. For example, techniques of observation from the study of operant conditioning offer a great deal to persons

concerned with measuring the effects of cueing procedures and reinforcement contingencies on response rate or frequency (See, for example, Honig, 1966; Baer, Wolf, and Risley, 1968; and Weiss, 1968). A major work on observation couched in the science of ethology is also a must for students of observational technique (Hutt and Hutt, 1970).

Methods for recording data obtained through observation include diary records, checklists, rating scales, rate and frequency counts, and anecdotal records. Details of these methods are given in many sources (e.g., Adams, 1964; Furth, 1958; Payne, 1968; Stanley, 1964), which should be consulted by readers unfamiliar with such methods. For the present, some basic features of observational systems suitable for use in early childhood education will be mentioned.

Traditionally, systems for observing young children have been focused on the individual child, his social and problem-solving skills, play activities and interations with materials in the classroom (Simon and Boyer, 1970). An example of a recently-developed system for individual child observation is the Personal Record of School Experience -- PROSE -- (Medley, 1969). This system involves no rating, but simply the objective recording of observable events as they occur. One child at a time is observed and all of his activity is recorded by means of a manageable coding system based on 11 categories of behavior (e.g., level of attention, manifest affect, and physical activity). Static conditions such as class organization, subject matter, and instructional materials in use can also be recorded. Codified data may be computer-analyzed. The PROSE is based upon the principle of OsCAR, a widely known system for observing teacher-pupil interaction (Medley, 1963).

Other examples of new observational schemes include systems for assessing preschool classroom environments (Stern and Gordon, 1967), nine aspects of young

children's classroom behavior (Katz, 1968), and social behavior in natural settings (Honig et al, 1970; Cunningham and Boger, 1971). The Stern and Gordon (1967) inventory of checklists and scales is notable for its comprehensiveness. Categories for measurement include (1) physical environment, materials and equipment, (2) program structure, balance, and organization, (3) play activities, (4) predominant teaching mode, (5) role of the teacher regarding verbal and nonverbal communication, (6) group control and management, (7) teacher involvement in children's social relations, (8) classroom atmosphere, (9) teacher "style and tone," and (10) general aspects of the teacher's relationship with children.

Systems for observing the relative strengths and abilities of teachers continue to be developed. A good example is Brown's (1970) Teacher Practices Observation Record. This instrument is being used in several Project Follow Through settings as an aid in developing effective behavior in both teachers and teacher aides. However, the most extensive developments in classroom observational technique are based on the concept of interaction analysis. The impetus for such developments has been largely provided by the Flanders' Interaction Analysis Technique (Flanders, 1966). The application of this technique requires that an observer keep a running record of teacher-pupil exchanges in three-second intervals. These exchanges can be tabulated according to categories of behavior that range from direct (e.g., giving commands, lecturing, justifying authority) to indirect teacher influences (e.g., accepting pupils' ideas and feelings, praising). Provision for recording extent of pupil talk is also made. Resulting data can be used to analyze prevailing patterns of teacher-student interaction and the relationship of such patterns to pupil achievement and attitudes.

Much research based on the Flanders system has accumulated in the past decade. For recent reviews, see Nuthall (1970) and Garrard (1966). Thus, if nothing else, this approach has been of great heuristic value. However, the

technique also has immense practical value for guiding teacher behavior. Mild criticism has been levelled at the Flanders system because of (1) its exclusive focus on verbal classroom behavior (especially of the teacher) and (2) its predominant concern with affective components of classroom behavior. But several extensions of this technique have been made which merit consideration by those involved in measuring and evaluating teacher behavior. These extensions include greater provision for cognitive factors (Amidon, 1966) (Reynolds et al, 1971) and nonverbal classroom interaction (French and Galloway, 1968). Perhaps the single most important finding from interaction analysis research for early childhood educators is that student teachers taught interaction-analysis are generally more indirect, supporting, and accepting of their pupils than are student teachers unfamiliar with this approach (Amidon, 1967).

Space limitations do not permit an elaborate review of all exciting developments in the area of classroom observation. However, a few additional systems deserve brief mention. These include systems designed especially for assessing student teachers (Sharpe, 1969), teacher skill in classroom management (Soar et al, 1971), a process approach to teacher's question-asking behavior (Zimmerman and Bergan, 1968), and the Behavioral Analysis Instrument for Teachers (1969). The latter is particularly useful for describing teacher skill in pedagogical technique, curriculum planning, and pupil evaluation-diagnosis. Perhaps the most comprehensive of all the newer systems is the Classroom Observation Instrument (SRI, 1970) developed for use in evaluating Project Head Start and Follow Through. Finally, interested readers are encouraged to examine carefully Simon and Boyer's (1970) anthology of 79 classroom observation systems, and two major publications about the role of systematic observation for assessing and improving classroom behavior (Brown, 1969; Gallagher, Nuthall, and Rosenshine, 1970).

There has been no extensive use of systematic observation techniques for measuring parental behavior in connection with early childhood education. There are at least two reasons for this. First, it is only recently that educators have become sensitive to the role of parents in the formal early education enterprise. Second, it is extremely difficult to arrange for such observation, either in homes or schools. It is therefore much more common for parental behavior to be measured by interviews, questionnaires, and checklists. However, the potential of observing parent-child interaction to measure such things as parental teaching styles, the quality of parent-child relationships, and home stimulation cannot be overlooked. The value of such an approach is well illustrated by the work of Hess and Shipman (1965), Bee et al (1969), Brophy (1970), Schmidt and Hore (1970), and STDP (1969). Observational procedures have also been utilized in evaluating maternal inservice training associated with preschool intervention programs (Hamilton, 1971).

A word about rating scale methods for tabulating observational data is also in order. It is clear from the literature that ratings of teachers by supervisors and of children by teachers continue to be popular and expedient means for measuring classroom behavior. For example, a recent survey of 53 of the nation's 60 largest school districts revealed that 50 of these districts currently use some type of rating scale to measure teacher performance (Queer, 1969). Problems and procedures associated with these and other techniques for measuring faculty instructional effectiveness are discussed by Blair (1968) and Cohen and Brawer (1969). Concerning effectiveness among teacher educators, it is not surprising that primary factors in "good" instruction include (1) coursework in which objectives are clearly defined, (2) a classroom atmosphere conductive to student ease, and (3) a tolerant, responsive instructor who

demonstrates both competence and enthusiasm (Bannister, 1961). Such characteristics would seem also to apply to teachers of young children. Research continues to highlight the importance of qualities such as empathy, nurturance, and communication skill for early childhood educators. For cues concerning the measurement of such qualities see Hogan (1969) and O'Leary and Becker (1969).

The problems of reliability and validity inherent in rating scale approaches to measurement are well known. But the use of well designed scales for measuring children's behavior by experienced teachers is often beneficial. Two examples illustrate such benefit. First, a rating scale adapted from the face sheet of the Stanford-Binet Intelligence Test form has been reported as being extremely useful in measuring three important motivational characteristics of children: Achievement motivation, confidence in ability, and activity level (Hess et al, 1966). Second, promising rating scale devices for predicting kindergarten and primary grade academic achievement, learning difficulties, and behavior problems have been reported (Attwell et al, 1967) (Conrad and Tobiessen, 1967) (Gross, 1970).

Rating scale methods also figure heavily in efforts to measure "socialization" of bilingual and ethnic group children (e.g., CerVenka, 1968), teachers' estimates of social competency among preschool and elementary school children (Levine and Elzey, 1968; Seagoe, 1970), and infant development (Hoopes, 1967).

Standardized, Objective Measures. This class of measures includes instruments for the measurement of intelligence and aptitude, achievement, personality, attitude, and interest. Such measures commonly appear in the form of tests that are constructed, administered, and scored according to prescribed rules (Brown, 1971). These rules govern the selection of item content, instructions for giving

and taking the test, and recording and evaluating test responses. Strictly speaking, such tests are limited to the measurement of behavior in the specific test situation, a situation which is usually contrived. Consequently, any statements or conclusions about the person being tested represent an inference from that situation to the general class of responses presumably sampled by the test. That is, by way of inference one generalizes from sample of test behavior to the broader characteristic(s) of the individual. This is particularly true of norm-referenced measures. Finally, it should be noted that not all tests are limited to a paper-and-pencil format, nor do all tests require formal arrangements. In this sense, any measure of performance can be called a "test." However, this section of the chapter is concerned largely with formal, obtrusive tests administered either individually or in groups.

With respect to standardized, objective measures of children's behavior, only a few points are made here. The reader is referred to the Head Start Test Collection (Educational Testing Service) for an annotated bibliography of commercially available tests and scales for use with children; and a handbook of such measures not commercially available has been published elsewhere (Johnson and Bommarito, 1971). First, conventional measures of mental ability continue to be used extensively for purposes of diagnosing developmental status, guidance, and measuring the effects of therapy (Stott and Ball, 1965). By far the most frequently used measure of young children's intelligence is the Stanford-Binet. Other widely used measures are Goodenough's Draw-a-Man, the Weschler Intelligence Scales for preprimary and school-age children, the Gesell Schedules, the Cattell Infant Scale, Ammons Picture-Vocabulary, and the Merrill-Palmer Scale. For complete reviews of these and other conventional scales see Stott and Ball (1965).

Second, alternatives to conventional measures of intelligence are finding

favor among many psychologists and educators (Achenbach, 1970). This is particularly true for those who have been attracted to Piaget's cognitive-developmental theory of mental development. Piaget-based scales for use as early as infancy have been developed (e.g., Uzgiris and Hunt, 1969; Honig and Lally, 1970). Other scales for the measurement of precausal thinking, object permanence, classificatory development, and conservation have been devised (Laurendeau and Pinard, 1962; Decarie, 1965; Kofsky, 1966; Goldschmid and Bentler, 1968). A critical examination of this developmental approach to the measurement of cognition and its implications for practitioners has been provided by Sullivan (1967).

Third, there has been a tremendous surge of interest in the measurement of children's language competence since the advent of federal compensatory education programs. Newly developed language measures are appearing regularly in the literature, many of which are used to measure the outcomes of language training programs for disadvantaged and minority group children (e.g., Bierly, 1971; Mehrabian, 1970; Stern and Gupta, 1970). Further, research workers are beginning to explore how standardized testing procedures may be altered better to assess the language skills of disadvantaged preschoolers. For example, a modified Peabody Picture Vocabulary Test- has been devised whereby three important variables -- expectancy for success, reinforcement, and specificity of task instructions -- are accounted for in the test administration (Ali and Costello, 1971). The net effect of this modification has been positive in terms of enhanced test scores for preschool children who otherwise may respond less well under "conventionally standard" conditions.

Fourth, the influence of humanistic psychologies is apparent in the now widespread concern for children's affective development among early childhood educators. Unfortunately, the validity and other technical features of most

measures of children's affect are unimpressive, if not poor. According to Hoepfner (1970), few worthwhile measures of achievement motivation, interest, activity level, and self-esteem are available. Paradoxically, these are among the phenomena about which some educators are most concerned. In this writer's opinion, however, genuine attempts to develop better measures in these areas are becoming both more frequent and fruitful (e.g., Adkins, 1968; Bolea, 1970; and Soares and Soares, 1969).

Not surprisingly, a majority of these attempts have focused on self-concept measures. An annotated bibliography of currently available measures of this construct designed for use with young children can be obtained through the ERIC Clearinghouse on Early Childhood Education (Coller, 1970). Unfortunately, most of these measures are marked by serious limitations: They invite socially desirable responses, depend heavily on young children's verbal facility, and utilize terminology the meaning of which is subject to wide differences in interpretation. In view of such limitations, some research workers (e.g., Long and Henderson, 1970; Yeatts and Bentley, 1971) have experimented with a non-verbal approach to self-esteem with modestly encouraging results. Other pertinent resources relevant to measurement in the affective domain are Beatty (1969), Bloom et al (1971, Chapter 10), and Eiss and Harbeck (1969). The latter two sources in particular deal with the knotty problem of affective objectives.

Finally, it should be noted that, apart from experimental programs in early childhood education, the systematic use of standardized, objective measures by nursery and kindergarten teachers apparently is not extensive. For example, Goslin et al (1965) report little use of tests beyond reading readiness and individual intelligence tests at the kindergarten level. Gross IQ

data are of little use to teachers faced with the complexities of educational planning (Neisworth, 1969). Even the results of reading readiness tests, when obtained, are not often put to good use (Goslin, 1967). These practices are due to at least two reasons. One is the limited number of educationally useful tests available to teachers in the past. Fortunately, this state of affairs is rapidly changing. Another reason may be that teachers of young children simply are not trained to use such measures, including their selection, administration, and interpretation. In this writer's judgment, training along these lines is important at both the pre- and in-service teaching levels. Such training for both testing and systematic observation conceivably can promote greater teacher initiative, cooperation, and responsibility concerning classroom measurement practices. As teacher involvement increases it seems more likely that classroom measurements will be put appropriately to use. Certainly educators should not allow tests to be administered and interpreted by untrained personnel.

While the need for teacher skill in test selection, administration, and interpretation is critical, a precautionary word is in order. As Carroll (1970) has observed, standardized tests can be overused and too much reliance placed on their results. For Carroll (1970) the problem is twofold: First, a given standardized test may not be sufficiently appropriate to the particular learning tasks in a local curriculum; and, second, the overall score or grade level index derived from standardized test performance may be inadequate for determining what specific skills have and have not been well acquired by a student. These limitations must be kept in mind by practitioners who elect to use standardized tests.

Consider next the use of standardized, objective measures of teacher behavior. Like interview measures, tests and scales for the measurement of teacher

behavior are most frequently used to select and predict instructional effectiveness. Occasionally, they are used to assess the effects of teacher training. Regardless, their use is more extensive at the pre-service rather than the in-service level. Published tests that have received more than occasional use include the California Psychological Inventory (Gough, 1968), Minnesota Test of Teacher Attitudes (Yee and Fruchter, 1971), The Teacher Preference Schedule (Stern and Masling, 1958), the Watson-Glaser Test of Critical Thinking (Watson and Glaser, 1952), and the Tennessee Self-Concept Scale (Fitts, 1965). Still other measures of interest in the study of teacher behavior can be cited. One of more than average usefulness for prediction purposes is addressed to teacher's beliefs about learning and teaching and the effect of such beliefs on classroom atmosphere (Harvey et al, 1966). Measures of teacher knowledge and ability to apply principles of good teaching in simulated problem situations have been developed (Popham, 1964) (Murray, 1969); and a method of assessing teacher attitudes toward children's behavior problems is available (Tolor et al, 1967).

The value of measures such as these depends on the purpose for which they are being used. It appears that more mileage can be obtained by employing a systematic observational approach, especially if some indication of teaching effectiveness is sought. More frequently than not, a low and positive insignificant relationship is obtained between performance on paper-and-pencil tests and teacher behavior as perceived by disinterested classroom observers. The issue here concerns the degree of correspondence between observed skill and verbalized beliefs, attitudes, and professed knowledge about teaching. Among the more promising steps in the direction of measuring degree of correspondence between teacher intentions and actual practices has been taken by Steele (1969). The resultant technique appears useful in determining the extent to which an instructional treatment is stably executed.

The common validity problem of paper-and-pencil tests of teaching skill is in part responsible for the development of performance tests of teaching effectiveness. Micro-teaching is one example of a technique that can be used to obtain some measure of actual performance. Still other performance approaches to the measurement of teaching proficiency have been attempted (e.g., Popham, 1971; Moody and Bausell, 1971). However, such approaches have usually failed to differentiate experienced, formally trained teachers from inexperienced, non-teachers. Perhaps the performance tests are faulty, but it is possible that such results indicate the inadequate nature of many teacher education programs.

Finally, standardized, objective measures of student opinion, attitudes toward instruction, and achievement are being increasingly used as indications of teaching effectiveness. However, this occurs mainly at the college level wherein preservice teachers rate or otherwise evaluate their teaching faculty. As yet, little work has been done to develop measures of preschool or early school pupil reactions to teachers. Strickland's (1970) report of explorations with a school attitude questionnaire for young children is a notable exception. No attempt will be made here to review the vast literature of student evaluation of teaching. Sources of information about measurement in this area include Davidoff (1970), Evans (1969), Hayes (1968), Hoyt (1969), Justiz (1969), Lewis (1966), McKeachie (1969) and Paraskevopoulus (1968).

Parents, understandably, are not much tested in connection with early childhood education programs. When they are, it is usually in the form of scales to measure attitudes toward child rearing practices and education or perceptions of themselves and their children in relation to training objectives (e.g., IPLET, 1969). The measurement of parent attitudes has a long history which has involved the development of a variety of scales for research use, some of which conceivably could be put to good use by educators (Baumrind, 1967;

Lorr and Jenkins, 1963; Schaefer and Bell, 1958). Yet most measures of this kind are beset with problems of both validity and reliability. Caution in their application is therefore warranted. Measurement procedures that have been used in home teaching and parent involvement projects are discussed by Kemble (1969), Orhan and Radin (1969) and Weikart and Lambie (1968). Finally, a scale designed to assess parent attitude change in relation to community action programs has been devised by Hanson, Stern, and Kitana (1968).

Teacher-Made Tests. This category of measurements includes short answer, objectively scored tests, essay and written documents, and many pupil products (e.g., art work, written materials, constructions, and various classroom projects). Most readers are familiar with such measures. Their nature, construction, and use are described in any basic textbook about educational measurements. Comments about teacher-made tests are therefore limited here to three incidental points.

First, behavioral objectives in any program of instruction in effect can themselves become measurements of the criterion-reference type. That is, if one describes (1) an individual's behavior that is to be performed together with (2) the context conditions of performance specifically enough so that the behavior can (3) be recognized when it occurs, then one's measurement task is straightforward: Observe and record the behavior. However, it is usually necessary also to specify a desired minimal level of performance (Mager, 1962). Such a suggestion is especially appropriate for those teachers who design their instructional programs around a mastery concept of achievement (See Block, 1971).

Second, in this writer's judgment, the potential of pupil products for measuring developmental progress, including academics, is frequently underestimated. However, there is some indication that a pupil product orientation is preferred even to conventional testing by many educators, especially those who

identify with "open education."[3] For example, such educators maintain that the "best measure of a child's work is his work" (Barth, 1969). Any meaningful application of this principle obviously requires that careful records of children's work be kept. An analysis of the cumulative change in children's work on a longitudinal basis is also necessary. Admittedly, relatively "informal" measurements are extremely limited for research purposes. But children's conceptual functioning, problem solving skills, and aesthetic expression can all be revealed in unique ways by activities that result in pupil products of various kinds.

Third, and finally, there is a great need for improvement in the test-making skills of teachers in early childhood education programs. Too often, teacher education programs require no course at all or require only a general course about tests and measurements in which descriptive statistics and item writing are stressed. Since the principal focus of such courses usually is on tests that require literacy, prospective early childhood personnel often see them as irrelevant. In short, more attention is needed to the development of skills in constructing checklists, tests of sensory discrimination and vocabulary, procedures for evaluating pupil products, and possibly even interview measurements among prospective preprimary teachers. Combined lack of skill in measurement technique and lack of understanding of how measurements can be used to facilitate instruction may also explain why teachers of young children often fail to incorporate a measurement perspective into their educational programs.

Fortunately, useful resources specifically concerned with the construction of informal measurement procedures are beginning to appear. For example,

---

[3] Ostensibly, the open education movement represents more a commitment to the process of learning, including the enhancement of cognitive processes. Ultimately, however, some evidence of process -- as reflected in the child's behavior (a product of some kind) -- is necessary for evidence of process.

assessment tools for teachers of pre-primary children who have various language and learning disorders are described by Bangs (1968). Bangs also provides curriculum suggestions once assessments are made. Another resource devoted to informal educational measurement has much to recommend it (Smith, 1969). This author deals with the areas of perceptual-motor development, reading and arithmetic skills, handwriting and spelling, speech and language disorders, and personal-social behavior.

Other Measurement Techniques. Thus far nothing has been said about projective measures, the measurement of social relations (sociometry), creativity measurement, and the medical approach to behavior measurement, including biological structure and function.[4] In a pragmatic sense, there is good reason for these "oversights." Projective measures of "personality," for example, are rarely used outside the clinical setting. Even many clinicians seemingly have become disenchanted with projective techniques because of their low validity. Moreover, teachers are not trained to administer and interpret projective measures; nor, in this writer's estimation, should they be. Readers interested in the use of projective techniques with children are referred to Levine (1966) and Blum (1968).

Sociometry has made a unique contribution to our understanding of social phenomena such as popularity and friendship, peer acceptance and rejection, leadership and influence power, group roles, and the relationship of sociability to school achievement. Insights concerning these phenomena have come mainly from the study of children beyond the preprimary level. But the successful use of sociometrics with nursery and kindergarten children has been reported (e.g., Northway, 1969a; 1969b; Hartup, 1970). Even so, sociometrics seem largely to be

---

[4] Measurement considerations associated with still another concept, cost-benefit analysis, are not dealt with in this paper. For an introduction to this approach see Alkin (1970).

utilized by child development research workers, not teachers and psychological specialists in the public schools. There are probably a number of reasons for this. One is lack of knowledge about and skill in using sociometric devices on the part of teachers. Another is that early childhood programs, while ostensibly devoted to promoting children's social development, infrequently reflect specific goals that call for systematic measurement in this area. Still another is the occasional ethical objection raised in connection with sociometrics, that is, a reluctance to "meddle" in children's social lives. The irony of this should be self-evident. Regardless, in this writer's estimation, the potential of sociometrics for gaining a better understanding of children's social perceptions, competence, and acceptance has not been much capitalized upon by educators. Again, interested readers are referred to other sources for a more comprehensive treatment of sociometric theory and technique (Gronlund, 1959; Northway and Weld, 1957).

Creativity is a much discussed, but little understood characteristic of human behavior. Not surprisingly, most empirical approaches to the measurement of creativity lack both a consensus about the behavior being measured and technical refinement (Tryk, 1968). Despite the measurement controversy, one cannot help being impressed with the vast amount of creativity research that has accumulated in recent years.[5] Most such research has been with older children, youth, and adults. More recently, however, considerable effort has been deployed to measure young children's creative potential under experimental conditions. For example, behavioral tasks deemed relevant to such potential have been devised by Starkweather (1966). These tasks are purported to measure psychological freedom,

---

[5] For a recent and comprehensive review of creativity research see Wallach (1970).

willingness to try difficult tasks, curiosity, and originality. Ideational fluency, rate, uniqueness, expressive freedom, productivity, and communicability are criteria variously stressed in still other recent attempts to measure creativity in early childhood (Ward, 1969a; ward, 1969b; Gross and Marsh, 1970; Singer and Whiton, 1971). Finally, it is claimed that certain portions of the Torrance Tests of Creative Thinking are appropriate for children as young as age four (Torrance, 1966). For an overview of problems involved in measuring young children's creativity, see Starkweather, 1964).

Early childhood personnel concerned with the assessment of physical growth, physiological functioning, sensory awareness, and the like should consult the following sources for pertinent surveys of measurement technique and research methodology: Eichorn (1970), Kaye (1970), Macy and Kelly (1960), Meredith (1960), Reisen (1960), and Tanner (1970). Diverse approaches to the measurement of temperament persistence, curiosity, impulse control, and reflectivity also merit the attention of educators. Not only are these measurable characteristics related to young children's school achievement, but they possibly can aid as indicators of affective development (Banta, 1970; Kagan, 1965; Maccoby et al, 1965; Maw and Maw, 1970; Thomas, Chess, and Birch, 1971). Finally, the Illinois Test of Psycholinguistic Abilities (1970) should be noted for its use as a frame of reference in diagnostic-prescriptive teaching.

The When of Measurement

Thus far, the what and how of measurement have occupied most of our attention. The question of when to measure theoretically should be answered according to one's evaluation plan. In the previous chapter, much was said about formative and summative evaluation. Both types of evaluation obviously call for measurement of one sort or another. There is little reason to elaborate further on this matter except a reminder that, in educational practice, most analyses require that measurements be taken at least two points in time. Too often, measurement occurs only in the context of summative evaluation with little attempt to assess entering or baseline behavior.

The issue of immediate and long-term measurement of behavior also is applicable to the when question. As far as most training programs are concerned, measurement is usually limited to immediate outcomes--performance during and/or

at the end of a given program (Webb, 1970). This emphasis is usually accompanied by the tacit assumption that performance at such times is an appropriate, if not good, indicator of how well one will perform over the longer haul. If the skills being learned during teacher-training, for example, are directly linked to eventual classroom teaching performance, then the assumption is valid. Otherwise, the measurement of performance during training may only be remotely related to measurement of on-the-job proficiency.

In the case of both teachers in training and children in early education programs, a basic question is whether short-term measurements involve anything more than behavior developed over the short term. Related to this is the problem of determining how soon one can expect program effects to be realized; and to this problem can be added that of determining with certainty how stable and durable are the behavior changes brought about by programs (Caro, 1971). The ideal procedure to follow would involve repeated, if not continuous, measurement of program "output variables" (Caro, 1971). Unfortunately, one-time measurement in connection with immediate-term summative evaluation is probably more the rule than the exception in actual practice. However, systematic follow-up measurement is a sound and potentially enlightening policy for both teacher educators and teachers of young children.

SELECTING AND EVALUATING MEASUREMENT TECHNIQUES

In the earlier discussion of criterion- and norm-referenced measurement, the point was made that decisions about measurement strategy should be made on the basis of what decisions will be made from the measurements obtained. This decision-making process will necessitate a careful consideration of program objectives. The role of objectives is also instrumental to the selection of measurement

techniques. Broadly speaking, there are at least two ways in which this selection can be made. First, measurement techniques can be determined in advance directly from specific program objectives. If such objectives are unique, then measurement techniques often must be designed locally. That is, existing tests, scales, or other techniques simply may not be valid or otherwise suitable for the measurement of one's instructional objectives.

Second, one can approach this selection problem only with broad program goals in mind, then select from "off the shelf" some existing measurement technique(s) for his purposes. If so, the measurements obtained become either (1) *ipso facto* instructional objectives and/or (2) a sample of behavior that may be spuriously related to the content and objectives of one's curriculum.

The first procedure is generally preferable to the second, simply because it demands clear, systematic thinking about the exact purposes of a given instructional program. However, as suggested earlier in the chapter, measurement need not be limited to a narrow set of instructional objectives. One is often interested in determining change or outcomes for which specific curriculum experiences were not arranged. For example, a teacher might be specifically concerned with measuring pupil progress in language skill development brought about through pattern drill. But this same teacher may also wish to obtain a measure of the extent to which children become more or less anxious about school during language instruction, even though the curriculum *per se* has no formal provision for modifying anxiety level. In fact, this notion is reflected in the thinking of others (e.g., Caro, 1971) who recommend that any possible unintended program effects, including those which are undesirable, should be both anticipated and measured.

Regardless, the issue of selecting measurement techniques in relation to program objectives is essentially a matter of validity: Will a given technique

measure what I want to measure? While validity is unquestionably the most critical aspect of any measurement endeavor, other qualities of measurement techniques are also important. Unless a given technique can measure behavior reliably or consistently from one time to the next, the interpretation of behavior changes in relation to an instructional program is next to impossible. Even the most valid, reliable measures may not be feasible for use in some programs due to complexities in administration, scoring, cost, or other factors. Thus, practicality is still another important consideration. Finally, in the case of norm-referenced measurements, the quality of norming procedures clearly is crucial, especially when important decisions about individuals are to be made. In short, the criteria of validity, reliability, and practicality are basic to any process of selecting and evaluating measurement techniques. Readers who wish more information about these criteria and their application to technique selection and evaluation tasks are referred to Brown (1970), Cronbach (1970), and the American Psychological Association publication concerning test standards (1966).

SPECIAL CONSIDERATIONS IN MEASUREMENT

Despite obvious values, the process of measurement is marked by many problems and issues that demand the attention of educators. These problems and issues can be grouped into three related categories: (1) The problem of academic or cultural bias, (2) the general impact of testing on students, and (3) how test results are used (Bloom, 1969) (Goslin, 1968). In this section, these related groupings of problems and issues in measurement are examined. In this way, the necessity for an ethical approach to educational measurement can be better clarified.

Academic or Cultural Bias

Most tests that are used with children of late preschool age and beyond already contain "academic" components which require specific language, discrimination, and conceptual skills (space, number, time) for successful performance. These tests also frequently call for various amounts of scholastic information (Stephens and Evans, in press). For purposes of assessing achievement status and predicting subsequent scholastic success, this academic orientation is suitable *if* two conditions can be met. First, a child must have encountered the *opportunity* to become familiar with the test-related content in a general way. Second, a child must have at least a minimal repertoire of test-taking skills. Many young children obviously can meet neither condition. Therefore, a test with a strong academic flavor may be a very poor sample of such children's past learning.

Admittedly, a tester can do little about a child's limited familiarity with culturally- or academically-based test content except, of course, temper his interpretation of a child's performance accordingly and avoid mis-using the test results. In other words, the problem is not so much a matter of tests *per se*. Rather, it is more a problem of the test user.

The problem of test-taking skills, however, is a matter about which a tester can do something more concrete. Personnel involved in test administration should provide for children experiences that simulate test conditions prior to the formal test itself. This includes provisions for practice in following directions, handling test materials, and the like. A "warm-up" period prior to testing is also advisable. An example of this policy in action can be cited. Oakland and Weilart (1971) implemented special activities for disadvantaged children to develop their familiarity with tests and skills necessary for test

performance. These activities had an initially strong, positive effect on the standardized test performance of these children as compared to peers who received no special training.

Again, the problem of academic or cultural bias is basically one of test validity. This problem of test validity often is most apparent when tests are given to make decisions about children whose performance may be affected negatively by factors such as low reading ability, low test-taking motivation, lack of familiarity with the language and conceptual style of a test, negative attitudes toward school personnel and academics, and poorly designed tests formats or instructions (Freeburg, 1970). These factors frequently are noted when conventional tests are used with children from impoverished circumstances or certain ethnic and minority groups.

The argument that tests based on standard English discriminate negatively and unfairly against children whose native language or dialect differs from the standard form has recently gained much support. This argument underlies, in part, the development of experimental measures of language competence and translations of conventional tests (e.g., the Stanford Binet Intelligence Scale) into non-standard English forms. As far as variant English dialects are concerned, the overall results of such efforts are mixed. For example, some authorities (e.g., Baratz, 1969) report salutary effects when tests are administered and scored in terms of the child's native dialect. This is notably true in the case of measuring native dialect proficiency. Similarly, others (e.g., Garvey and McFarlane, 1970) have traced variations in standard English proficiency among black children from lower socioeconomic homes to interference from their normal language pattern, rather than academic ability differences.

In contrast, still other authorities (e.g., Quay, 1971) report that intelligence test performance among black children is little affected by the language of test administration, i.e., whether the test is given in standard English or black dialect; and, no reliable performance differential has been observed among black children to whom a test of echoic responding was administered in either standard English or black dialect (Stern and Gupta, 1970).

Of course, there are at least two basic problems inherent in studies designed to compare children's performance on different linguistic forms of the same test. One is that tests translated from standard English to a second language (e.g., Spanish or French) may lose some validity in the very process of translation--to say nothing of the questionable practice of using norms from the original version of the test in order to describe performance on the translated version. A second problem is that no one English dialect is likely to characterize all children from different ethnic or racial minorities. Thus, comparisons of standard English with black dialect, for example, may be gross and misleading.

Clearly, extensive research is needed to clarify the influence of dialect on test-taking behavior. But certainly tests based on standard English are unsuitable for use with any child who neither understands nor can speak the language of the test. This means that a teacher must first ensure that a child is sufficiently competent in standard English if such competence is necessary for a valid test performance. Otherwise, alternative methods of measurement must be sought in order to obtain any reasonably accurate portrait of a child's achievement status. In short, the practice of early classifying or otherwise evaluating with standard English measures the intellectual competence of children whose native language is different from English cannot be condoned. For an example of early intervention research in which a bilingual approach to measurement is illustrated, see Nedler and Sebere (1971).

General Impact of Testing on Students

As noted earlier in this chapter, a given testing program may have as its purpose nothing more than obtaining useful information. However limited the intended purposes of a testing program, the effect of testing will doubtlessly extend beyond this point (Stephens and Evans, in press). Simply taking a test, or expecting one for that matter, is bound to have various effects on most children (Goodwin, 1966). Unfortunately, controlled research in this area of psychology has been meagre. But reasoned speculation combined with what little research has been done in this area can lead to the identification of some definite possibilities. For example, testing effects may occur in advance of actual testing by influencing the type and degree of preparation in which students engage themselves. In some cases, teachers (and parents) may even coach their charges both in the tactics of test-taking and the content of anticipated test situations.

Effects may also be realized during a test itself. For example, testing can act as a form of teaching (Stephens and Evans, in press). In taking a test early in an educational program, students may learn something about what they will be studying (or evaluated on later) and also become sensitized to the material that will be stressed. Students may thus be led to pay more attention to this material when it is encountered (Entwhistle, 1961).

While this may often be desirable from a teacher's point of view (particularly in the case of criterion-referenced measurement), some disadvantages may occasionally occur from testing. Some students, for example, in advancing an erroneous answer, may become more committed to that answer. Subsequently, they may encounter some difficulty in overcoming the misconception or inaccuracy that their answer represents (Stephens and Evans, in press). Finally, every precaution should be taken to avoid conveying the message to students that test-taking and strong test performance are the end-alls of education.

As Bloom (1969) suggests, the psychological effects on students during an actual examination may be comparatively light, except for such possibilities as excessive anxiety or emotional stress, frustration, self-doubt, or feeling of failure (or accomplishment!) that may be associated with a test situation. Further, while fatigue at the end of an extended examination may occur among some children, it is likely that this effect is usually mild and short-lived.

But perhaps the most potentially serious outcome in this regard can come from "overtesting" children in connection with early intervention programs. It has been this writer's experience that some evaluators simply schedule too much formal testing during the course of an early education program. Conceivably, the net adverse effect of this problem could be at least threefold. First, children may come to view testing negatively because it can mean unwelcome intrusions and undue pressures for performance. Second, teacher resentment over

interruptions of their instructional program can accumulate to the point where their cooperation with testers may suffer. Third, overtesting is subject to the law of diminishing returns: That is, the time and expense involved in extensive testing may not be worth the additional results obtained. It is also possible that "too much" data can serve to complicate, rather than facilitate, the interpretation of program outcomes. This concern over excessive testing was also discussed in the first evaluation chapter.

There is no universal rule for determining how many of what kind of tests one should utilize in an evaluation plan. Nevertheless, a judicious perspective on this issue is imperative.

It has been suggested that temporary fatigue or inconvenience occasionally associated with testing are comparatively minor problems. However, prolonged anxiety that may be engendered by testing and evaluation procedures is not. Among other things, the literature concerned with test anxiety indicates that teachers (testers) should avoid adding emotionalism to testing procedures by dramatizing the hazards of doing poorly or the idea that a student's future is at stake in the testing situation. It is therefore extremely important to acknowledge the potentially debilitating effects that intense emotion may have on an individual's test performance, particularly where complex intellectual tasks are involved. There is good evidence to indicate that individual differences in test anxiety are apparent among children as early as kindergarten entrance and that subsequent changes in measured anxiety level are linked to patterns of change in achievement and intelligence test performance throughout the elementary school period; in general, it is the highly anxious child or youth that stands to suffer the most in this regard (Sarason et al, 1960) (Sarason, Hill and Zimbardo, 1964).

Thus far, a few ideas concerning the impact of testing on students before and during examination periods have been presented. While these dimensions of the problem are important, postexamination effects possibly are the most profound, depending upon what uses are made of test results. As Bloom (1969) has observed, these effects may be minimal or maximal, positive or negative, but they can neither be completely controlled nor entirely neutralized.'

Post-testing effects are in part a function of the type of test utilized and the way in which tests are utilized. For example, there are at least three categories of tests that carry a high potential for lasting effects: (1) Tests designed to measure significant and relatively stable human qualities such as tests of intelligence and aptitudes; (2) tests that are used to facilitate major educational decisions -- for example, tests for admission to certain academic programs, certification of satisfactory completion of an educational program, and the like; and (3) tests whose results become a permanent part of a student's record or that are made public for one reason or another (Bloom, 1969). Extreme care must therefore be exercised in regard to the selection, administration, and interpretation of tests used for such purposes. This leads to an explicit ethical consideration of the way in which test results may be used.

The Ethical Use of Test Results

Concern about the possible misuses of test results is represented by an extensive literature (e.g., Black, 1963; Dyer, 1961; Ebel, 1964; Hoffman, 1962; Mehrens and Lehmann, 1969). Among the more critical potential misuses of tests that are discussed in this literature are four summarized by Ebel (1964). These misuses can serve both to illustrate the current and widespread professional concern about tests and suggest to the reader some guidelines for his own policy formulation about tests.

First, it is conceivable that imprudent educational testing can indelibly mark a student's intellectual status as superior, average, or inferior. If so, his subsequent academic or social status could be more or less predetermined by way of expectancies that become established among those privy to test results and decisions about educational programming that come about through testing. This need not necessarily be destructive to the individual. But an individual who is "assigned" a label of "weak student" on the basis of, say, an intelligence test score, may be adversely affected both in self-esteem and motivation for future achievement.

Second, it is possible that certain testing practices can generate a restricted concept of human abilities, one based largely on degrees of success in intellectual achievement situations. Consequently, this sort of concept may lead to a focus upon the attainment of limited goals, often at the expense of educational practices that are designed to facilitate the development of diverse talent.

A third possible misuse of test results concerns the exercise of Machiavellian tendencies among those in charge of testing programs. By this is meant the exercise of excessive and unwarranted control over the personal destinies of children.

Finally, poorly conceived testing practices may foster rigid, mechanistic, and depersonalized approaches to measurement and evaluation that, in effect, could limit basic human rights and impede positive human relations within the schools.

Such distasteful outcomes are not inevitable. However, it is clear that steps must be taken to guarantee that these outcomes do not materialize. Tests should be viewed as but one of several means for increasing student achievement

by way of motivating and directing the energies of students and teachers alike. Furthermore, the use of tests to impose upon others certain decisions and courses of action should be deferred as much as possible to their use in providing data for choice in individual decision making (Ebel, 1964). As previously indicated, the issue therefore concerns more the way tests are used than it does the nature of tests per se.

This writer has often been impressed by the negative emotionalism associated with tests and their use by many parents, college students, and teachers. Frequently this emotionalism leads to exaggerated claims about the inhumane or even subversive nature of testing, particularly intelligence and personality testing. Such emotionalism undoubtedly is kindled by inadequate understanding of tests and their uses. But, instances of unwise test use in the school can provide justification for much criticism. At the extreme, tests may be confiscated and burned by opponents of psychometrics (Nettles, 1959; Eron and Walden, 1961). Less extreme, but indicative of resistance, is refusal by parents to submit their children to testing in the school setting. Still another area of conflict for educators is the matter of when, by what means, and how extensively parents should be informed about their children's test performances. As a matter of course, professionals who are responsible for specific programs of childhood education and research must develop a judicious policy in relation to these problems. At the very least, this policy should include advance parental permission for testing and an acceptable method for communicating the purposes and outcomes to concerned parents. This recommendation of course, is based on the assumption that parents have the right of access to their children's school records.

Unfortunately, virtually no empirical data exist concerning the effects on parents of receiving information about their children's test performances (Kirkland, 1971). Research in this area is sorely needed in order to glean better clues for policy formulation about school testing practices.[6]

It is also pertinent to consider the social consequences of abandoning tests. In a broad sense, the case for no testing is much the same as that advanced against evaluation in general. [redacted] It is unnecessary to elaborate in great detail on this issue. However, the following quote summarizes well a reasoned position on this matter:

> "If the educational tests were abandoned, the encouragement and reward of individual efforts,would be made more difficult. Excellence in programs of education would become less tangible as a goal and less demonstrable as an attainment. Educational opportunities would be extended less on the basis of aptitude and merit and more on the basis of ancestry and influence; social class barriers would become less permeable. Decisions on important issues of curriculum and method would be made less on the basis of solid evidence and more on the basis of prejudice or caprice. These....are likely to be the more harmful consequences, by far. Let us not forego the wise use of tests." (Ebel, 1964, p.334).

Implicit in this passage is the notion that problems of interpreting and using test results may occur largely because of certain lacks--for example, lack of knowledge about the limitations of tests and the technical and theoretical aspects of testing. Yet if test results are to be useful they must be communicated to those directly responsible for students, singly or in groups (Levine, 1966). Again, the task here is one of establishing sound policy for communication, a policy that includes safeguard for student welfare.

This section of the chapter can be concluded with a reference to a position statement on psychological assessment recently adopted by the American Psychological Association (1970). This statement in fact represents a policy for

---

6 Readers interested in a further study of testing policy are referred to Vol. 20, No. 11 of the American Psychologist (1965). The entire issue is devoted to issues and ethics associated with testing and public policy.

testing and the use of test results formulated in relation to the essential features of psychological assessment. As such, it is an extremely important policy for consideration by school personnel everywhere.

1. Guaranteed protection must be provided for every individual against unwarranted inferences by educational personnel ill-equipped with necessary background knowledge and skill in testing.

2. Obsolete information (dated test results and the like) that might lead to unfavorable evaluation of an individual must be periodically culled from personal records in order to protect that individual.

3. Unnecessary intrusions into one's privacy must be avoided; irrelevant tests and questions have no place in a well-designed assessment program.

4. Given the above modes of protection, procedures should be established to facilitate continual investigation of new and improved techniques of assessment.

While these guidelines are pertinent to ability testing, they are perhaps even more applicable to personality assessment. In either case, the key concept again is relevance. That is, measurement procedures should have demonstrable relevance to the peculiar purposes of evaluation -- whether one wishes to evaluate academic competence, instructional effectiveness, or personal-social development. However, one must simultaneously determine relevance in relation to an ethical framework and criteria of social acceptability.

SUMMARY AND CONCLUSIONS

This essay represents a selective overview of contemporary measurement practices in early childhood education. Although techniques for the measurement

of young children's behavior received primary attention, comment was extended to procedures for measuring both teacher and parent behavior.

A number of emergent trends concerning current measurement practice were cited. Those with perhaps the deepest implications for early education reflect attempts to broaden the spectrum of measurement in several directions, including a greater sensitivity to cultural or ethnic interests and a stronger focus upon children's affective development. However, much technical work remains to be done before field application of measures relevant to these foci can confidently be made.

Measurement was considered along three basic lines of thought: what, how, and when. An answer to the question of what to measure requires a careful scrutiny of educational objectives. In turn, the specification of objectives is instrumental in determining an answer to the how of measurement. Achievement of a consistency among program objectives, instructional content, and measurement procedures is imperative for sound curriculum evaluation.

It is clear that interviews, systematic observational procedures, and tests (standardized and otherwise) continue to dominate educational practice. However, many creative variations of these techniques have recently appeared. Moreover, a re-birth of mastery approaches to learning is apparent from the widespread interest in criterion-referenced measurement. Measurement is also being applied increasingly to variables other than strict behavioral outcomes of educational programs. These variables include curriculum components, curriculum materials, teacher-child interaction, and the physical environment for education. Among other things, this means that comprehensive measurement will include the input and process phases of instruction, as well as the traditional output phase. A precise conceptualization of basic input, process, and output variables can also serve better to frame the question of when to measure.

Finally, the research literature reviewed herein suggests that direct behavioral measures are generally preferable to those that only provide data for inferences about hypothetical constructs, primarily for reasons of validity. Validity is also a central issue in the special problems involved when many conventional measures are used with individuals from varying cultural backgrounds. Cultural bias, along with the possible psychological effects of testing on children and the ethical use of test results, were identified as phenomena that warrant the attention of teachers and research workers alike.

Measurement Practices in Early Childhood Education: Bibliography

E. D. Evans, University of Washington

Achenbach, T. M. "The children's associative responding test: a possible alternative to group IQ tests." Journal of Educational Psychology, 1970, 61(5), 340-348.

Adams, Georgia Sachs. Measurement and evaluation in education, psychology, and guidance. New York: Holt, Rinehart and Winston, 1964.

Adkins, Dorothy. Measurement of motivation to achieve in preschool children. ERIC: ED 021 617, 1968.

Ahr, A. E. "The development of a group preschool screening test of early school entrance potentiality." Psychology in the Schools, 1967, 4, 59-63.

Ali, Faizunisa and Costello, Joan. "Modification of the Peabody Picture Vocabulary Test." Developmental Psychology, 1971, 5, 86-91.

Alkin, M. C. "Evaluating the cost-effectiveness of instructional programs." In M. C. Wittrock and D. E. Wiley (Eds.) The evaluation of instruction: issues and problems. New York: Holt, Rinehart and Winston, 1970, 221-238.

American Psychological Association. Standards for educational and psychological tests and manuals. Washington, D. C.: American Psychological Association, 1966.

American Psychological Association. "Psychological assessment and public policy." American Psychologist, 1970, 25, 264-267.

Amidon, E. "Interaction analysis: Recent developments." Paper read at the American Educational Research Association Meeting, Chicago, February, 1966.

Amidon, Edmund. The effect upon the behavior and attitudes of student teachers of training cooperating teachers and student teachers in the use of interaction analysis as a classroom observational technique. ERIC: ED 021 777, 1967.

Astin, A. W. and Holland, J. L. "The environmental assessment technique: a way to measure college environments." Journal of Educational Psychology, 1961, 52, 308-316.

Attwell, A. A., Orpet, R. E., and Meyers, C. E. "Kindergarten behavior ratings as a predictor of academic achievement." Journal of School Psychology, 1967, 6, 43-46.

Baer, D., Wolf, M., and Risley, T. "Some current dimensions of applied behavior analysis." Journal of Applied Behavior Analysis, 1968, 1, 91-97.

Bangs, Tina E. Language and learning disorders of the pre-academic child. New York: Appleton-Century-Crofts, 1968.

Bannister, J. Evaluating college teaching. ERIC: ED 022 450, 1961.

Baratz, Joan C. "A bi-dialectical task for determining language proficiency in economically disadvantaged Negro children." Child Development, 1969, 40, 889-901.

Barth, R. S. "Open education." Theory and Philosophy of Education, Kensington, Australia, 1969.

Baumrind, Diana. Parent attitude inquiry: Research edition III. Berkeley, Calif.: Department of Psychology, University of California, 1967.

Baumrind, Diana. Naturalistic observation in the study of parent-child interaction. ERIC: ED 027 073, 1968.

Beatty, W. H. Improving educational assessment and an inventory of measures of affective behavior. Washington, D. C.: Association for Supervision and Curriculum Development, NEA, 1969.

Bee, Helen et al. "Social class differences in maternal teaching strategies and speech patterns." Developmental Psychology, 1969, 1, 726-734.

Behavioral analysis instrument for teachers. Syracuse, N. Y.: Educational and Cultural Center, 1969.

Bierly, Margaret M. "A validation of a method of assessing young children's language competence." Paper read at the American Educational Research Association Meeting, New York, February, 1971.

Black, H. They shall not pass. New York: Morrow, 1963.

Blair, B. Evaluation of faculty instruction. ERIC: ED 029 642, 1968.

Block, J. H. (Ed.) Mastery Learning. New York: Holt, Rinehart and Winston, 1971.

Bloom, B. S. "Some theoretical issues relating to educational evaluation." In R. W. Tyler (Ed.) Educational evaluation: new roles, new means. Chicago: University of Chicago Press, 1969, 26-50.

Bloom, B. S., Hastings, J. T., and Madaus, G. F. Handbook of formative and summative evaluation of student learning. New York: McGraw-Hill, 1971.

Blum, G. S. "Assessment of psychodynamic variables by the Black Pictures." In P. McReynolds (Ed.) Advances in psychological assessment, Vol. I. Palo Alto, Calif.: Science and Behavior Books, 1968, 150-168.

Boehm, Ann. Test of basic concepts. New York: The Psychological Corporation, 1969.

Bolea, A. S. The development and validation of a pictorial self-concept scale for children in K-4. ERIC: ED 037 780, 1970.

Bormuth, J. R. (Ed.) Readability in 1968. Washington, D. C.: National Conference on Research in English, National Council of Teachers of English, 1968.

Brophy, J. E. "Mothers as teachers of their own preschool children: The influence of socioeconomic status and task structure on teaching specificity." Child Development, 1970, 41, 79-94.

Brown, B. B. (Ed.) Systematic observation: relating theory and practice in the classroom. ERIC: ED 031 444, 1969.

Brown, B. B. "Systematic observation and analysis of teaching." In Roberta LaCoste (Ed.) Focus on Kindergarten, Early Childhood Education Series. Olympia, Wash.: State Department of Public Instruction, 1970.

Brown, F. G. Principles of educational and psychological testing. Hinsdale, Ill.: The Dryden Press, 1970.

Butler, Annie L. Current research in early childhood education. Washington, D. C.: American Association of Elementary-Kindergarten-Nursery Educators, 1970.

Caldwell, Bettye M. The preschool inventory, Princeton, N. J.: Educational Testing Service, 1967.

Caro, F. G. "Issues in the evaluation of social programs." Review of Educational Research, 1971, 41, 87-114.

Carroll, J. B. "Problems of measurement related to the concept of learning for mastery." Educational Horizons, 1970, 43, 71-80.

Cazden, Courteney B. "Evaluation of learning in preschool education: Early language development." In B. Bloom, J. T. Hastings, and G. F. Madaus (Eds.) Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971, 345-398.

Cervenka, E. J. Administration manual for the inventory of socialization of bilingual children ages three to ten. ERIC: ED 027 062, 1968.

Cohen, A. M. and Brawer, Florence B. Measuring faculty performance. Washington, D. C.: American Association of Junior Colleges, 1969.

Coller, A. R. An annotated bibliography of self-concept measures for the early childhood years. Urbana, Ill.: ERIC Clearinghouse on Early Childhood Education, 1970.

Conrad, W. G. and Tobiessen, J. "The development of kindergarten behavior rating scales for the prediction of learning and behavior disorders." Psychology in the Schools, 1967, 4, 359-363.

Creager, J. A. and Astin, A. W. "Alternative methods of describing characteristics of colleges and universities." Educational and Psychological Measurement, 1968, 28, 719-734.

Cronbach, L. J. Essentials of psychological testing (3rd Ed.). New York: Harper and Row, 1970.

CSE. Elementary school test evaluations. Los Angeles: Center for the Study of Evaluation, UCLA, 1970.

Cunningham, Jo Lynn and Roger R. P. "Development of an observational rating schedule for preschool children's peer group behavior." Paper read at the American Educational Research Association Meeting, New York, February, 1971.

Davidoff, S. H. The development of an instrument designed to secure student assessment of teaching behaviors that correlate with objective measures of student achievement. Philadelphia: Office of Research and Evaluation, Philadelphia Public Schools, 1970.

Deal, T. and Wood, P. "Testing the early educational and psychological development of children." Review of Educational Research, 1968, 38, 12-18.

Decarie, T. T. Intelligence and affectivity in early childhood. New York: International University Press, 1965.

Dick, W. "A methodology for the formative evaluation of instructional materials." Journal of Educational Measurement, 1968, 5, 99-102.

DuBois, P. H. and Mayo, C. D. (Eds.). Research strategies for evaluating training. AERA Monograph Series on Curriculum Evaluation. Chicago: Rand McNalley, 1970.

Dyer, H. S. "Is testing a menace to education?" New York State Education, 1961, 49, 16-19.

Eash, M. J. "Assessing curriculum materials: a preliminary instrument." Educational Product Report 2, No. 5, 1969, 18-24.

Ebel, R. L. "The social consequences of educational testing." School and Society, 1964, Nov., 331-334.

Ebel, R. L. Some limitations of criterion-referenced measurement: ERIC: ED 038 670, 1970.

Eichorn, Dorothy. "Physiological development." In P. Mussen (Ed.) Carmichael's manual of child psychology. New York: Wiley, 1970, 157-283.

Eisa, A. F. and Harbeck, Mary Blatt. Behavioral objectives in the affective domain. Washington, D. C.: National Science Supervisors Association, 1969.

English, H. B., and English, Ava C. A comprehensive dictionary of psychological and psychoanalytic terms. New York: David McKay, 1958.

Engelmann, S. Basic concept inventory. Chicago: Follett, 1967.

Entwhistle, Doris. "Attensity: Factors of specific set in school learning." Harvard Educational Review, 1961, 31, 84-101.

Evans, E. D. "Student activism and teaching effectiveness: survival of the fittest?" Journal of College Student Personnel, 1969, 10, 102-108.

Fitts, W. H. Tennessee self concept scale. Nashville: Counselor Recordings and Tests, 1965.

Flanders, N. A. Interaction analysis in the classroom (Rev. Ed.). Ann Arbor, Mich.: School of Education, University of Michigan, 1966.

Freeburg, N. E. "Assessment of disadvantaged adolescents: A different approach to research and evaluation measures." Journal of Educational Psychology, 1970, 61, 229-240.

French, R. L. and Galloway, C. M. A description of teacher behavior: verbal and nonverbal. ERIC: 028 134, 1968.

Furth, E. J. Constructing evaluation instruments. New York: Longmans, Green and Co., 1958.

Gallagher, J. J., Nuthall, G. A., and Rosenshine, B. Classroom observation. AERA Monogrpah Series on Curriculum Evaluation. Chicago: Rand McNalley, 1970.

Garrard, Judy. Classroom interaction: review of the literature. ERIC: ED 013 988, 1966.

Garvin, A. D. The applicability of criterion-referenced measurement by content area and level. ERIC: ED 041 038, 1970.

Glaser, R. "Instructional technology and the measurement of learning outcomes: Some questions." American Psychologist, 1963, 18, 519-521.

Goldschmid, M. L. and Bentler, P. M. "The dimensions and measurement of conservation." Child Development, 1968, 39, 787-802.

Goodwin, W. L. "Effect of selected methodological conditions of dependent measures taken after classroom experimentations." Journal of Educational Psychology, 1966, 57, 350-358.

Goolsby, T. M. and Darby, Barbara M. A bibliography of instrumentation methodology and measurement in early childhood learning. Athens, Ga.: University of Georgia, Research and Development Center in Educational Stimulation, 1969.

Goslin, D. A. "Standardized ability tests and testing." Science, 1968, 159, 851-855.

Goslin, D. Teachers and testing. New York: Russell Sage, 1967.

Goslin, D., Epstein, R., and Hallock, B. The use of standardized tests in elementary schools. Technical Report No. 2: The Social Consequences of Testing. New York: Russell Sage, 1965.

Gough, H. G. "An interpreter's syllabus for the California Psychological Inventory." In P. McReynolds (Ed.) Advances in psychological assessment (Vol. I). Palo Alto, Calif.: Science and Behavior Books, 1968, 55-79.

Gronlund, N. E. Sociometry in the classroom. New York: Harper, 1959.

Gross, M. B. "Preschool prediction of academic achievement." Psychological Reports, 1970, 26, 278.

Gross, Ruth B. and Marsh, Marion. "An instrument for measuring creativity in young children: The gross geometric forms." Developmental Psychology, 1970, 3, 267.

Guilford, J. P. Psychometric methods (2nd Ed.). New York: McGraw-Hill, 1954.

Hamilton, M. I. Evaluation of a parent and child center program. ERIC: 045 189, 1971.

Hanson, Susan A., Stern, Carolyn, and Kitana, H. L. Attitude differences related to economic status. Los Angeles: Early Childhood Research, 1968.

Hartup, W. W. "Peer interaction and social organization." In P. Mussen (Ed.) Carmichael's manual of child psychology (Vol. II) (3rd Ed.) New York: Wiley, 1970, 361-456.

Harvey, O. J., White, B. J., Prather, M. S., Alter, R. D., and Hoffmeister, J. K. "Teachers' belief systems and preschool atmospheres." Journal of Educational Psychology, 1966, 57, 373-381.

Hayes, R. B. A way to evaluate and to improve classroom teaching effectiveness. ERIC: ED 048 095, 1968.

Headley, Neith. The kindergarten: its place in the program of education. New York: Center for Applied Research in Education, 1965.

Hess, R. et al. Techniques for assessing cognitive and social abilities of children and parents in Project Head Start. ERIC: ED 015 772, 1966.

Hess, R. D. and Shipman, Virginia C. "Early experience and the socialization of cognitive modes in children." Child Development, 1965, 36, 869-886.

Hoepfner, R. Measuring normal affective states in children. Los Angeles: UCLA, Center for the Study of Evaluation, 1970.

Hoffman, B. The tyranny of testing. New York: Crowell, 1962.

Hogan, R. "Development of an empathy scale." Journal of Consulting and Clinical Psychology, 1969, 33, 307-316.

Holtzman, W. H. "The changing world of mental measurement and its social significance." American Psychologist, 1971, 26, 546-553.

Honig, Alice S., Caldwell, Bettye M., and Tannenbaum, J. "Patterns of information processing used by and with young children in a nursery school setting." Child Development, 1970, 41, 1045-1065.

Honig, Alice S. and Lally, J. R. Piagetian infancy scales. Syracuse, N. Y.: Syracuse University Children's Center, 1970.

Honig, W. K. Operant behavior: areas of research and application. New York: Appleton-Century-Crofts, 1966.

Hoopes, Janet L. An infant rating scale: its validation and usefulness. New York: Child Welfare League of America, Inc., 1967.

Hoyt, D. P. Identifying effective teaching behavior. ERIC: ED 039 197, 1969.

Hutt, S. J. and Hutt, Corrine. Direct observation and measurement of behavior. Springfield, Ill.: Charles C. Thomas, 1970.

IDS. Development of a standardized telephone interview for the measurement of language changes in young children. ERIC: ED 022 758, 1968.

IPLET. The implicit parental learning theory scale. Syracuse, N. Y.: Syracuse University Children's Center, 1969.

Jenkins, W. O. et. al. The status of behavioral measurement and assessment in children. ERIC: ED 018 252, 1966.

Johnson, O. G. and Bommarito, J. Tests and measurements in child development: A handbook. San Francisco: Jossey-Bass, 1971.

Justiz, T. B. "A reliable measure of teacher effectiveness." Educational Leadership, 1969, 27, 49-55.

Kamii, Constance. "Evaluation of learning in preschool education: Socio-emotional, perceptual-motor, cognitive development." In B. Bloom, J. T. Hastings, and G. F. Madaus (Eds.). Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971, 281-344.

Kamii, Constance and Elliot, D. L. "Evaluation of evaluations." Educational Leadership, 1971, 28, 827-831.

Kamii, Constance and Peper, R. A Piagetian method of evaluating preschool children's development in classification. ERIC: ED 039 013, 1969.

Katz, Lillian G. Child behavior survey instrument: Manual of instruction and definition. Stanford, Calif.: Stanford University Center for Research and Development in Teaching, 1968.

Kaye, H. "Sensory processes." In H. W. Reese and L. P. Lipsitt (Eds.) Experimental Child Psychology, New York: Academic Press, 1970, 33-64.

Kemble, Virginia. A feasibility study of parent awareness programs. Final report. ERIC: ED 040 742, 1969.

Kim, Y., Anderson, H. E., Jr., and Bashaw, W. L. "The simple structure of social maturity at the second grade level." Educational and Psychological Measurement, 1968, 28, 145-153.

Kirschner Associates, Inc. A national survey of the impacts of Head Start centers on community institutions. Washington, D. C.: Office of Child Development, U. S. Office of Health, Education, and Welfare, 1970.

Klare, G. R. The measurement of readability. Ames, Iowa: Iowa State University Press, 1963.

Kofsky, Ellin. "A scalogram study of classificatory development." Child Development, 1966, 37, 191-204.

Kritchevsky, Sybil and Prescott, Elizabeth. Planning environments for young children: physical space. Washington, D. C.: National Association for the Education of Young Children, 1969.

Laurandeau, M. and Pinard, A. Causal thinking in the child. New York: International Universities Press, 1962.

Lavatelli, Celia Stendler. Piaget's theory applied to an early education curriculum. Boston: American Science and Engineering, 1970.

Lay, Margaret and Dopyera, J. Analysis of early childhood programs: A search for comparative dimensions. Urbana, Ill.: ERIC Clearinghouse on Early Childhood Education, 1971.

Levine, M. "Psychological testing of children." In M. L. Hoffman and Lois W. Hoffman (Eds.). Review of child development research, Vol. II. New York: Russell Sage, 1966, 257-310.

Levine, S. and Elzey, F. F. Development of a social competency scale for preschool children. ERIC: ED 020 004, 1968.

Lewis, Gertrude M. The evaluation of teaching. ERIC: ED 030 602, 1966.

Long, Barbara H. and Henderson, E. H. "Social schemata of school beginners: Some demographic correlates." Merrill-Palmer Quarterly, 1970, 16, 305-324.

Lorr, M. and Jenkins, R. "Three factors in parent behavior." Journal of Consulting Psychology, 1963, 17, 306-308.

McKeachie, W. J. "Student ratings of faculty." American Association of University Professors Bulletin, 1969, 55, 439-444.

McReynolds, P. "An introduction to psychological assessment." In P. McReynolds (Ed.) Advances in psychological assessment (Vol. I). Palo Alto, Calif.: Science and Behavior Books, 1968, 1-13.

Macy, Icie G. and Kelly, Harriet J. "Chemical and physiologic growth." In P. Mussen (Ed.). Handbook of research methods in child development. New York: Wiley, 1960, 252-283.

Mager, R. F. Preparing instructional objectives. Palo Alto, Calif.: Fearon, 1962.

Medley, D. M. "Experiences with the Oscar technique." Journal of Teacher Education, 1963, 14, 267-273.

Medley, D. M. Oscar goes to nursery school: A new technique for recording pupil behavior. Princeton, N. J.: Educational Testing Service, 1969.

Mehrabian, A. "Measures of vocabulary and grammatical skills for children up to age six." Developmental Psychology, 1970, 2, 439-446.

Mehrens, W. A. and Lehmann, I. J. Standardized tests in education. New York: Holt, Rinehart, and Winston, 1969.

Meredith, H. V. "Methods of studying physical growth." In P. Mussen (Ed.) Handbook of research methods in child development. New York: Wiley, 1960, 201-251.

MGS. The measure of a good school: A guide to evaluation of school systems adapted particularly for use in Kentucky School districts. ERIC: ED 037 920, 1964.

Moody, W. B. and Bauswell, R. B. "The effect of teacher experience on student achievement, transfer, and retention." Paper read at the American Educational Research Association Meeting, New York, February, 1971.

Moss, M. H. Tests of basic experiences. Monterey, Calif.: CTB/McGraw-Hill, 1970.

Murray, C. K. "The construct validity of the Teaching Situation Reaction Test." Journal of Educational Research, 1969, 62, 323-328.

Nedler, Shari and Sebera, Peggy. "Intervention strategies for Spanish-speaking children." Child Development, 1971, 42, 259-267.

Neisworth, J. T. "The educational irrelevance of intelligence." In R. M. Smith (Ed.). Teacher diagnosis of educational difficulties. Columbus, Ohio: Charles Merrill, 1969, 30-46.

Northway, Mary L. "The stability of young children's social relations." Educational Research, 1969, 11, 54-57. (a)

Northway, Mary L. "The changing pattern of young children's social relations." Educational Research, 1969, 11, 212-214. (b).

Northway, Mary L. and Weld, L. Sociometric testing: A guide for teachers. Toronto: University of Toronto Press, 1957.

Nunnally, J. C. Educational measurement and evaluation. New York: McGraw-Hill, 1964.

Nuthall, G. A. "A review of some selected recent studies of classroom interaction and teaching behavior." In J. J. Gallagher, G. A. Nuthall, and B. Rosenshine (Eds.) Classroom Observation. AERA Monograph Series on Curriculum Evaluation. Chicago: Rand McNalley, 1970, 6-29.

Oakland, T. and Weilert, Evelyn. "The effects of test-wiseness materials on standardized test performance of preschool disadvantaged children." Paper given at the American Educational Research Association Meeting, New York City, February, 1971.

O'Leary, K. D. and Becker, W. C. "The effects of the intensity of a teacher's reprimands on children's behavior." Journal of School Psychology, 1968-69, 7, 8-11.

Orhan, Shije and Radin, Norma. "Teaching mothers to teach: A home counseling program for low income parents." The Family Coordinator, 1969, 18, 326-311.

Painter, Genevieve. Infant education. San Rafael, Calif.: Dimensions Publishing, 1968.

Palmer, J. O. The psychological assessment of children. New York: Wiley, 1970.

Paraskevopoulos, I. "How students rate their teachers." Journal of Educational Research, 1968, 62, 25-29.

Payne, D. A. The specification and measurement of learning outcomes. Waltham, Mass.: Blaisdell, 1968.

Popham, W. J. "Predicting student teachers' instructional behavior from a structured and unstructured test of professional knowledge." California Journal of Educational Research, 1965, 16, 7-13.

Popham, W. J. "Performance tests of teaching proficiency: rationale, development and validation." American Educational Research Journal. 1971, 8, 105-117.

Popham, W. J. and Husek, T. R. "Implications of criterion-referenced measurement." Journal of Educational Measurement, 1969, 6, 1-9.

Prescott, Elizabeth and Jones, Elizabeth. Group day care as a childrearing environment: an observational study of day care programs. Pasadena, Calif.: Pacific Oaks College, 1967.

Quay, Lorene C. "Language dialect, reinforcement, and the intelligence test performance of Negor children." Child Development, 1971, 42, 5-15.

Queer, G. An analysis of teacher rating scales: a national survey. ERIC: ED 030 968, 1969.

Rea, R. E. and Rays, R. E. "The comprehensive mathematics inventory: an experimental instrument for assessing youngsters entering school." Journal of Educational Measurement, 1970, 7, 45-47.

Reisen, A. H. "Receptor functions." In P. Mussen (Ed.) Handbook of research methods in child development. New York: Wiley, 1960, 284-310.

Reynolds, W. R., Abraham, E. C., and Nelson, M. A. "The classroom observational record." Paper read at the American Educational Research Association Meeting, New York, February, 1971.

Rogolsky, Maryrose M. "Screening kindergarten children: a review and recommendations." Journal of School Psychology, 1968-69, 7, 18-27.

Sapir, Selma G. and Wilson, Bernice. "A developmental scale to assist in the prevention of learning disability." Educational and Psychological Measurement, 1967, 27, 1061-1068.

Sarason, S. B. et al. Anxiety in elementary children. New York: Wiley, 1968.

Sarason, S. B., Hill, K. T., and Zimbardo, P. G. "A longitudinal study of the relation of test anxiety to performance on intelligence and achievement tests." Monographs of the Society for Research in Child Development, 1964, 29, Ser. No. 98, 51 pp.

Savage, R. D. Psychological assessment of the individual child. Baltimore: Penguin Books, 1968.

Schaefer, E. S. and Bell, R. Q. "Development of a parental attitude research instrument." Child Development, 1958, 29, 336-361.

Seagoe, May V. "An instrument for the analysis of children's play as an index of degree of socialization" Journal of School Psychology, 1970, 8, 139-145.

Schmidt, W.H.O., and Hore, T. "Some nonverbal aspects of communication between mother and preschool child." Child Development, 1970, 41, 889-896.

Shalock, H. D. Appendix Z: Classes of measures used in the behavioral sciences, the nature of the data that derive from them, and some comments as to the advantages and disadvantages of each. ERIC: ED 026 331, 1968

Sharpe, D. M. Isolating relevant variables in student teaching assessment. ERIC: 028 999, 1969.

Simon, Anita and Boyer, E. G. (Eds.) Mirrors for behavior II (Vols. A and B): An anthology of observation instruments. Philadelphia: Research for Better Schools, Inc., 1970.

Singer, D. L. and Whiton, Mary Beth. "Ideational creativity and expressive aspects of human figure drawing in kindergarten age children." Developmental Psychology, 1971, 4, 366-369.

Sjogren, D. D. "Measurement techniques in evaluation." Review of Educational Research, 1970, 40, 301-320.

Smith, R. M. (Ed.) Teacher diagnosis of educational difficulties. Columbus, Ohio: Charles Merrill, 1969.

Soar, R. S., Ragosta, Marjorie, and Soar, Ruth M. "The validation of an observation system for classroom management." Paper read at the American Educational Research Association Meeting, New York, Fevruary, 1971.

Soares, A. T. and Soares, Louise M. "Self-perceptions of culturally disadvantaged children." American Educational Research Journal, 1969, 6, 31-45.

SRI. Classroom observation instrument. Palo Alto, Calif.: Stanford Research Institute, 1970.

Stanley, J. C. Measurement in today's schools (4th Ed.). Englewood Cliffs, N. J.: Prentice-Hall, 1964.

Starkweather, Elizabeth. "Problems in the measurement of creativity in preschool children." Journal of Educational Measurement, 1964, 1, 109-113.

Starkweather, Elizabeth K. Potential creative ability and the preschool child. ERIC: ED 018 900, 1966.

Stedman, D. J. An approach to the study of infant behavior. ERIC: ED 039 031, 1966.

Steele, J. M. Things as they are: an evaluation procedure to assess intent and practice in instruction. Urbana, Ill.: University of Illinois, Unpublished Dissertation, 1969.

Stephens, J. M. and Evans, E. D. The psychology of classroom learning (2nd Ed.) New York: Holt, Rinehart and Winston, In press.

Stern, Carolyn and Gupta, Willa. "Echoic responding of disadvantaged preschool children as a function of type of speech modeled." Journal of School Psychology, 1970, 8, 24-27.

Stern, G. G. and Masling, J. M. Unconscious factors in career motivation for teaching. Washington, D. C.: U. S. Department of Health, Education, and Welfare, Office of Education, Contract No. SAE 6459, 1958.

Stern, Virginia and Gordon, Anne. Head Start evaluation and research center, progress report, document 4: Development of observation procedures for assessing preschool classroom environment. ERIC: ED 021 026, 1967.

Stevens, S. S. "Mathematics, measurement, and psychophysics." In S. S. Stevens (Ed.) Handbook of experimental psychology. New York: Wiley, 1951, 1-49.

STIM. Inventory of home stimulation. Syracuse, N. Y.: Syracuse University Children's Center, 1969.

Stott, L. H. and Ball, Rachell S. "Infant and preschool mental tests: review and evaluation" Monographs of the Society for Research in Child Development, 1965, 30, Ser. No. 101.

Sullivan, E. V. Piaget and the school curriculum: A critical appraisal. Toronto: Ontario Institute for Studies in Education, University of Ontario, Bulletin No. 2, 1967.

Tanner, J. M. "Physical growth." In P. Mussen (Ed.) Carmichael's manual of child psychology. New York: Wiley, 1970, 77-156.

Thomas, H. "Psychological assessment instruments for use with human infants." Merrill-Palmer Quarterly, 1970, 16, 179-223.

Tolor, A., Scarpetti, W. L., and Lane, P. A. "Teacher's attitudes toward children's behavior revisited." Journal of Educational Psychology, 1967, 58, 175-180.

Torrance, E. P. Torrance Tests of Creative Thinking. Princeton: N. J.: Personnel Press, 1966.

Tryk, H. E. "Assessment in the study of creativity." In P. McReynolds (Ed.) Advances in psychological assessment, Vol. I. Palo Alto, Calif.: Science and Behavior Books, 1968, 34-54.

Uzgiris, Ina C. and Hunt, J. McV. Toward ordinal scales of psychological development in infancy. In preparation, 1969.

Wallach, M. A. "Creativity." In P. Mussen (Ed.) Carmichael's manual of child psychology (3rd. Ed.), Vol. I. New York: Wiley, 1970, 1211-1272.

Ward, W. C. "Creativity and environmental cues in nursery school children." Developmental Psychology, 1969, 1, 543-547 (a).

Ward, W. C. "Rate and uniqueness in children's creative responding." Child Development, 1969, 40, 869-878 (b).

Watson, G. and Glaser, E. M. Manual, Watson-Glaser critical thinking appraisal. Yonkers-on-Hudson, N. Y.: World Book, 1952.

Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L. Unobtrusive measures: nonreactive research in the social sciences. Chicago: Rand McNalley, 1966.

Webb, W. B. "Measurement of learning in extensive training programs." In P. H. DuBois and C. D. Mayo (Eds.) Research strategies for evaluating training. Chicago: Rand McNalley, 55-65.

Weikart, D. P. and Lambie, Delores Z. "Preschool intervention through a home teaching program." In J. Hellmuth (Ed.) Disadvantaged child (Vol. II). Seattle: Special Child Publications, 1968, 435-500.

Weiss, R. L. "Operant conditioning techniques in psychological assessment." In P. McReynolds (Ed.) Advances in psychological assessment (Vol. I). Palo Alto, Calif.: Science and Behavior Books, 1968, 169-190.

White, B. W. Human infants. New York: Prentice-Hall, 1971.

Williams, J. and Fox, A. M. "Prediction of performance in student teaching." Educational and Psychological Measurement, 1967, 27, 1169-1170.

Yeatts, Pearline P. and Bentley, E. L. "The development of a non-verbal measure to assess the self-concept of young and low verbal children." Paper read at the American Educational Research Association Meeting, New York, February, 1971.

Yee, A. H. and Fruchter, B. "Factor content of the Minnesota Teacher Attitude Inventory." American Educational Research Journal, 1971, 8, 119-133.

Zimmerman, B. J. and Bergan, J. R. Intellectual operations in teacher-child interaction. ERIC: ED 039 011, 1968.

Ellis D. Evans

Additions to Bibliography: "Measurement Practices in Early Childhood Education"

Allport, G. W., Vernon, P. E., and Lindzey, G. A study of values. Boston: Houghton, 1951.

Banta, T. J. "Tests for the evaluation of early childhood education: The Cincinnati autonomy test battery." In J. Hellmuth (Ed.) Cognitive studies, Vol. 1. New York: Brunner-Mazel, 1970, 424-490.

Coopersmith, S. The antecedents of self esteem. San Francisco: Freeman, 1967.

Crandall, Virginia C., Katkovsky, W., and Crandall, V. J. "Children's beliefs in their own control of reinforcements in intellectual-achievement situations." Child Development, 1965, 36, 91-109.

Kren, L. and Walder, L. O. "Test burning: II." American Psychologist, 1961, 16, 237-244.

Harvey, Catherine, and McFarlane, P. "A measure of standard English proficiency of inner-city children." American Educational Research Journal, 1970, 7, 29-40.

Heckhausen, H. The anatomy of achievement motivation. New York: Academic Press, 1967.

Illinois Test of Psycholinguistic Abilities (Rev. Ed.). Urbana, Ill.: University of Illinois Press, 1970.

Kagan, J. "Reflection-impulsivity and reading ability in primary grade children." Child Development, 1965, 36, 609-628.

Kirkland, Marjorie C. "The effects of tests on students and schools." Review of Educational Research, 1971, 41, 303-350.

Lindvall, C. M. and Cox, R. C. The IPI evaluating program. AERA Monograph Series on Curriculum Evaluation. Chicago: Rand McNalley, 1970.

Maccoby, Eleanor E., Dowley, Edith M., Hagen, J. W., and Degerman, R. "Activity level and intellectual functioning in normal preschool children." Child Development, 1965, 36, 761-770.

Maw, W. H. and Maw, Ethel W. "Self concepts of high-and low-curiosity boys." Child Development, 1970, 41, 123-129.

Nettler, G. "Test burning in Texas." American Psychologist, 1959, 14, 682-683.

Sheldon, Eleanor and Moore, W. Indicators of social change. New York: Russell Sage, 1968.

Straus, M. A. Family measurement techniques. Minneapolis: University of Minnesota Press, 1969.

Strickland, G. Development of a school attitude questionnaire for young children. Los Angeles: UCLA, Center for the Study of Evaluation, 1970.

Teachers College. Indicators of Quality: A brochure. New York: Institute of Administrative Research, Teachers College, Columbia University, 1968.

Thomas, A., Chess, Stella, and Birch, H. G. Behavioral individuality in early childhood. New York: New York University Press, 1971.